Article Information

Mini Review

Deep Semantic Segmentation New Model of Natural and Medical Images

Pei-Yu Chen[1]*, Chien-Chieh Huang[2] and Yuan-Chen Liu[2]

[1]Department of Science Education, College of Science, National Taipei University of Education, Taipei City 10671, Taiwan

[2]Department of Computer Science, College of Science, National Taipei University of Education, Taipei City 10671, Taiwan

*Correspondence: Pei-Yu Chen, Department of Science Education, College of Science, National Taipei University of Education, Taipei City 10671, Taiwan, Email: g110324003@grad.ntue.edu.tw

Abstract

Semantic segmentation is the most significant deep learning technology.

At present, automatic assisted driving (Autopilot) is widely used in real-time driving, but if there is a deviation in object detection in real vehicles, it can easily lead to misjudgment. Turning and even crashing can be quite dangerous. This paper seeks to propose a model for this problem to increase the accuracy of discrimination and improve security. It proposes a Convolutional Neural Network (CNN)+ Holistically-Nested Edge Detection (HED) combined with Spatial Pyramid Pooling (SPP). Traditionally, CNN is used to detect the shape of objects, and the edge may be ignored. Therefore, adding HED increases the robustness of the edge, and finally adds SPP to obtain modules of different sizes, and strengthen the detection of undetected objects. The research results are trained in the CityScapes street view data set. The accuracy of Class mIoU for small objects reaches 77.51%, and Category mIoU for large objects reaches 89.95%.

Introduction

Convolutional Neural Network (CNN) has been widely used in the image field in recent years. In the beginning, the last layer of CNN used a fully connected layer (Fully Connected Layer) to classify and predict the probability, and the final output was in the form of numbers. However, in Fully Convolutional Networks (FCN) after [1] was proposed, the last layer was changed to a convolutional layer, an image of any size as input, restored to the original image size, and color distribution was performed to distinguish each semantic label by color to achieve end-to-end training. The method has also been renamed Semantic Segmentation, it is currently widely used in autonomous driving (Autopilot), robot vision (Computer vision) or medical image segmentation (Medical Image Segmentation). It can not only be used in daily life, autonomous driving prevents fatigue driving, robots can assist people in simple but labor-intensive tasks in life, and can even be used in medicine to assist doctors in strengthening the diagnosis of symptoms to achieve early treatment, and can enhance the accuracy of semantic segmentation. The rate is quite helpful with today's technological advancements.

Since AlexNet [2] proposed an 8-layer architecture in the 2012, ImageNet image classification competition, the top-5 error rate of AlexNet was 10% lower than that of the previous year's champion, successfully proving that CNN's deeper architecture can Neural networks with higher visual recognition, usually with two or more hidden layers (Hidden Layer), began to be named deep neural networks (Deep Learning). Drawing on the deep model architecture helps to improve the recognition rate. In 2014, the University of Oxford proposed VGGNet [3] with 11, 13, 16, and 19 layers of architecture, with 16 layers being the most common, called VGG16. It also improved the large convolution kernel of AlexNet to repeatedly extract features with repeated and small convolutions and obtained second place in ImageNet that year, and first place in the same year was the model developed by Google, GoogLeNet [4], which proposed 19 layers and included 9 Inception Module small modules. The next year, the champion ResNet [5]

in 2015, the depth was increased to 152 layers, and the number of parameters and recognition was greatly optimized. At this point, deep convolutional networks were more widely displayed in the field of technology.

The current semantic segmentation is mainly used for automatic assisted driving (Autopilot). At present, the mainstream trams or new gasoline-electric trams on the market are already equipped with them and can be officially put on the road, so they can be seen in real life. At present, it mainly uses multiple lenses that are installed on the car to detect road conditions [6,7]. If there is an emergency, such as a person, dog, ball, and so on, appearing on the driving route, emergency detection can be provided to prevent collisions for safety, and because semantic segmentation is different from object detection, it segments the fine edges of objects, while object detection is to detect the entire object, and it also has good identification during parking to prevent collisions [8]. If the combined system with higher accuracy can even achieve the popularization of automatic parking in the future, it is also seen that it can be mounted on drones and robots [9]. If semantic segmentation can be strengthened, it is expected that more precise operations can be performed in the future, so the semantics will be improved. Segmentation accuracy can not only be used in multiple fields but also increase the convenience and safety of life.

Literature

In the past, in terms of semantic segmentation, after Alex Net laid the foundation for image segmentation, deep models such as VGG, AlexNet, GoogLeNet, and ResNet appeared one after another. Many models have also tried to deepen or widen the model. Although recently in recent years, due to the significant improvement in the computing power of GPU (Graphics Processing Unit) hardware, even Nvidia has launched a TPU (Tensor Processing Unit) that specializes in computing artificial neural networks. In practical applications, parameter optimization is also an important issue. Several typical models of deepening models appeared, such as the PSPNet proposed by Hengshuang Zhao, et al. [10], which also uses the architecture of the deep model, combined with SPP (Spatial Pyramid Pooling), and improves the accuracy of training CityScapes. After high improvement, Google Liang-Chieh Chen and others proposed DeepLabV1 and DeepLabV2 [11,12] based on this. They not only added CRFs (Conditional Random fields) to the deepened model to enhance edge detection but in order to restore more feature details, SPP [13] was also added to Atrous and transformed into ASPP (Atrous Spatial Pyramid Pooling) [14], mainly for downsampling (Downsampling) to carry out a wider range of features. Extraction, and finally adding ASPP, enhanced feature extraction for original size images, greatly increasing the accuracy.

In addition to adding the enhanced detection module, Konrad Heidler, et al. [15] added Unet to HED (Holistically-Nested Edge Detection) greatly improving terrain detection. In Nvidia Gated-SCNN [16], it is also mentioned that color and texture are not important features for semantic segmentation, but edges are an important feature. It can enhance the restoration of objects, so edge detection is equally important, and semantic segmentation can be added to enhance object detection.

Research methodology

In previous papers, improvements are usually made by deepening or adding modules, and edges are also very important for detection. There are also many papers that add edge features, but the process of adding edges easily increases the amount of calculation, and there is one more The computation will make the overall time longer, and the detection of distant views is usually not very accurate.

Therefore, we hope to add edge detection or other modules to increase the overall accuracy without increasing many parameters and calculations. We use dual modules of semantic segmentation and edge detection to extract features and use deep ResNet for segmentation. The residual block in ResNet can not only reduce the excessive increase of parameters, use a deep model to extract a large number of features, and increase the basic accuracy, but also superimpose the high-resolution and low-resolution scales in ResNet in the module to increase the accuracy, and use HED edges. Detection, because HED can extract multi-layer edges in CNN without significantly increasing parameters, it can also increase edge features and strengthen feature applications. Compared with previous papers, CRF or edge operations are removed, and SPP and ASPP are finally added, which can feature extraction performed at large, medium, and small resolutions to capture missing features and improve the recognition rate, which can also reduce training time and resources.

At the end of the model, 3 x 3 convolution is usually used for the final output, and the output result is convolved into the last 3 channels. However, due to the high complexity of the module, the probability of overfitting also increases significantly, so we do it twice at the end. Convolution, and add Drop out in the middle, allowing neurons to randomly drop 0.2 to reduce the probability of overfitting in overly complex models. The overall architecture is as follows (Figure 1).

The figure above shows the model architecture. The CNN in the upper module has designed an active module that can be freely inserted into the model architecture. It is not limited to the specified model so it can provide better choices when different objects need to be detected. Also, due to the HED, this feature changes the upper-layer model. One only needs to freely change the number of layers one wants, and one can also obtain the edge segmentation map so that in dual-module restoration one can significantly assist the recognition rate without spending too many parameters and memory.

Conclusion

With the assistance of deeper models and dual modules, the accuracy can be confirmed. Reducing jump connections can also reduce the number of parameters and training speed, and using the highest resolution and lowest score can also effectively assist restoration. Finally, the addition of SPP and ASPP, ASPP proves that in complex scenes, because dilated convolution expands the target, features can be extracted more effectively. Overall spatial pyramid pooling can extract more large, medium and small features, remove CRF or edge applications similar to RNN, and
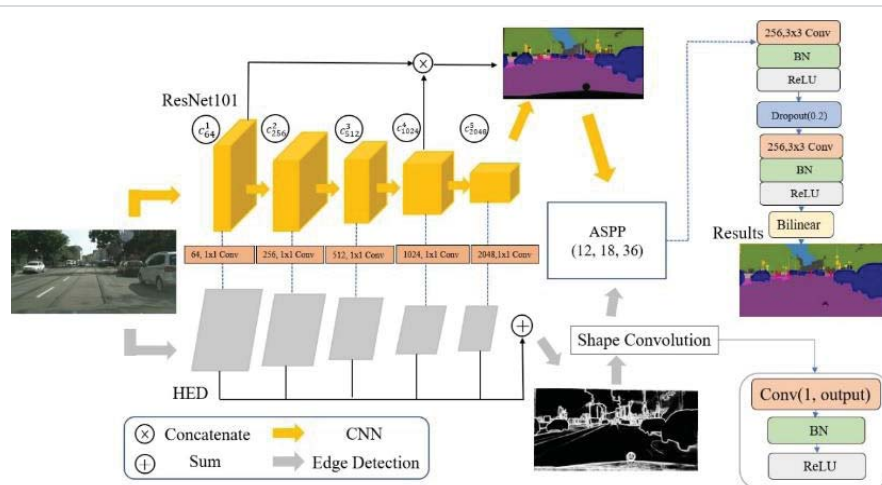
Figure 1: Model architecture diagram.

significantly reduce training. Parameters can also achieve higher accuracy, and the modules are also movable. If one has different training goals today, one can change the required model at any time, and the accuracy will not be lost due to high mobility.

The proposed model increases the accuracy of discrimination and improves security. It proposes CNN+HED combined with SPP. Adding HED increases the robustness of the edge, and finally, the addition of SPP to obtain modules of different sizes can strengthen the detection of undetected objects. The research results are trained in the CityScapes street view data set. The accuracy of Class mIoU for small objects has been found to reach 77.51%, and Category mIoU for large objects reaches 89.95%.

References

1. Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation. IEEE Trans Pattern Anal Mach Intell. 2017 Apr;39(4):640-651. doi: 10.1109/TPAMI.2016.2572683. Epub 2016 May 24. PMID: 27244717.

2. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems. Curran Associates Inc., Red Hook, NY, USA. 2012; 1:1097–1105.

3. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556. 2015.

4. Szegedy C. Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015; 1-9. doi: 10.1109/CVPR.2015.7298594.

5. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. arXiv:1512.03385. 2015.

6. Franke U. Making Bertha See, 2013 IEEE International Conference on Computer Vision Workshops. 2013; 214-221. doi: 10.1109/ICCVW.2013.36.

7. Cakir S, Gauß M, Häppeler K, Ounajjar Y, Heinle F, Marchthaler R. Semantic Segmentation for Autonomous Driving: Model Evaluation, Dataset Generation, Perspective Comparison, and Real-Time Capability. arXiv:2207.12939. 2022.

8. Hua M, Nan Y, Lian S. Small Obstacle Avoidance Based on RGB-D Semantic Segmentation. arXiv:1908.11675. 2019.

9. Girisha S, Manohara Pai MM, Verma U, Radhika M Pai. Semantic Segmentation of UAV Videos based on Temporal Smoothness in Conditional Random Fields. 2020 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER). 2020; 241-245. doi: 10.1109 /DISCOVER50404. 2020.9278040.

10. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid Scene Parsing Network. arXiv:1612.01105. 2017.

11. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans Pattern Anal Mach Intell. 2018 Apr;40(4):834-848. doi: 10.1109/TPAMI.2017.2699184. Epub 2017 Apr 27. PMID: 28463186.

12. Chen LC, Papandreou G, Schroff F, Adam H. Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv:1706.05587. 2017.

13. He K, Zhang X, Ren S, Sun J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. arXiv:1406.4729. 2015.

14. Chen LC, Papandreou G, Schroff F, Adam H. Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv. 2017.

15. Heidler K, Mou L, Baumhoer C, Dietz A, Zhu XX. HED-UNet: Combined Segmentation and Edge Detection for Monitoring the Antarctic Coastline. arXiv:2103.01849. 2021.

16. Takikawa T, Acuna D, Jampani V, Fidler S. Gated-SCNN: Gated Shape CNNs for Semantic Segmentation. arXiv:1907.05740. 2019.

INSTRUCTIONS FOR AUTHORS

IgMin Research - A BioMed & Engineering Open Access Journal is a prestigious multidisciplinary journal committed to the advancement of research and knowledge in the expansive domains of Biology, Medicine, and Engineering. With a strong emphasis on scholarly excellence, our journal serves as a platform for scientists, researchers, and scholars to disseminate their groundbreaking findings and contribute to the ever-evolving landscape of Biology, Medicine and Engineering disciplines.

For book and educational material reviews, send them to IgMin Research, at support@igminresearch.us. The Copyright Clearance Centre's Rights link program manages article permission requests via the journal's website (https://www.igminresearch.com). Inquiries about Rights link can be directed to info@igminresearch.us or by calling +1 (860) 967-3839.

https://www.igminresearch.com/pages/publish-now/author-guidelines

APC

In addressing Article Processing Charges (APCs), IgMin Research: recognizes their significance in facilitating open access and global collaboration. The APC structure is designed for affordability and transparency, reflecting the commitment to breaking financial barriers and making scientific research accessible to all.

At IgMin Research - A BioMed & Engineering Open Access Journal, fosters cross-disciplinary communication and collaboration, aiming to address global challenges. Authors gain increased exposure and readership, connecting with researchers from various disciplines. The commitment to open access ensures global availability of published research. Join IgMin Research - A BioMed & Engineering Open Access Journal at the forefront of scientific progress.

https://www.igminresearch.com/pages/publish-now/apc

WHY WITH US

IgMin Research | A BioMed & Engineering Open Access Journal employs a rigorous peer-review process, ensuring the publication of high-quality research spanning STEM disciplines. The journal offers a global platform for researchers to share groundbreaking findings, promoting scientific advancement.

JOURNAL INFORMATION

Journal Full Title: IgMin Research-A BioMed & Engineering Open Access Journal	**License:** Open Access by **IgMin Research** is licensed under a Creative Commons Attribution 4.0 International License. Based on a work at IgMin **Publications Inc.**
Journal NLM Abbreviation: IgMin Res	
Journal Website Link: https://www.igminresearch.com	
Topics Summation: 150	**Online Manuscript Submission:**
Subject Areas: Biology, Engineering, Medicine and General Science	https://www.igminresearch.com/submission or can be mailed to submission@igminresearch.us
Organized by: IgMin Publications Inc.	
Regularity: Monthly	
Review Type: Double Blind	
Publication Time: 14 Days	
GoogleScholar: https://www.igminresearch.com/gs	
Plagiarism software: iThenticate	
Language: English	
Collecting capability: Worldwide	